Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations 1. Thesis and Dissertation Collection, all items

1947

A Rating Scale for Marine Non-Commissioned Officers at the First Three Pay Grades

Baze, Grant Sheridan

The Ohio State University

http://hdl.handle.net/10945/46656

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

Downloaded from NPS Archive: Calhoun

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

A RATING SCALE FOR MARINE NON-COMMISSIONED OFFICERS OF THE FIRST THREE PAY GRADES

Grant Sheridan Baze

Thesis
B25

Thesis
B25

#2

A RATING SCALE FOR RATING NON-COMMISSIONED OFFICERS OF THE FIRST THREE PAY GRADES.

A Thesis Presented for the
Degree of Master of Science
in Public Administration.

By

Grant Sheridan Baze, B.B.A.

THE OHIO STATE UNIVERSITY

1947

Approved by:

ACKNOWLEDGMENTS

The writer expresses acknowledgement and appreciation to Professors Harvey Walker and J. E. Horrocks, who read and criticized the manuscript; and to Professor H. A. Edgerton, who checked the statistical principles involved.

He is indebted to the group of Naval officers pursuing a course in Personnel Administration and Training at Ohio State University for their valuable criticism and aid in constructing the scale: Captain A. B. Roby, Commander Lindsay Williamson, Lt. Commander R. F. Smith, Lt. Commander C. J. Walline, and Lt. (J.G.) A. H. Cornell. The same aid was contributed by Captain T. J. Mathews, U.S.M.C., Lt. (J.G.) David Chester, U.S.N.R. and Staff Sergeant C. A. Harrison, U.S.M.C.

To the Marine officers who answered the writer's questionaire, to the Army R.O.T.C. unit at Ohio State University for its cooperation in weighting the scale, and to the Naval R.O.T.C. unit for the extension of its facilities are due special thanks.

To his wife, Helen E. Bane, the writer is most indebted for her constant aid in preparation of the manuscript.

6352

TABLE OF CONTENTS

PAGE

I. INTRODUCTION: 1

II. GENERAL CONSIDERATIONS IN THE CONSTRUCTION OF THE RATING SCALE: 2

- A. Preliminary Considerations. 2
- B. Types of Scales Considered. 4
- C. Synthesis of some of the Principles of Constructing a Graphic Rating Scale. 11
- D. Some Methods of Validating Rating Scales. 16
- E. Some Methods of Establishing the Reliability of Rating Scales. 19
- F. Considerations in Instructing Raters. 19

III. THE MECHANICS OF CONSTRUCTING THE GRAPHIC RATING SCALE. 24

- A. A Letter Questionaire. 24
- B. Tabulation of the Suggested Characteristics. 27
- C. Definition of the Characteristics and Choice of Descriptive Phrases. 32
- D. Weighting the Items in the Scale. 41
- E. The Form of the Rating Sheet. 44
- F. FORMULATION OF INSTRUCTIONS TO ACCOMPANY THE SCALE 45
- F. Testing the Reliability of the Scale. 45

IV. SOME METHODS OF COMPENSATING FOR VARIATIONS IN RATERS' STANDARDS. 53

- A. The Method of Pooled Ratings. 53
- B. The Method of Standard Scores. 54
- C. The Use of Letter Grades. 55

PAGE

V. A SCHEME FOR THE USE OF THE SCALE: 57

A. Incorporating the scale into the Total Fitness Report. 57

B. Correcting Variations between Raters. 58

VI. CONCLUSIONS. 60

APPENDICES:

1. Instructions to Raters. 63

2. The Rating Sheets. 67

CHAPTER I

INTRODUCTION

During the latter part of 1946, instructions were issued by the Commandant, U. S. Marine Corps, concerning the use and marking of the fitness report for non-commissioned officers of the U. S. Marine Corps. Some loss of effectiveness had been experienced, due in part perhaps to the inexperience of the rating officers, but at any rate the ratings assigned appeared not to differentiate sufficiently between the various levels of qualification and performance. There was little spread in the distribution of markings.[1] Instructions were issued which will tend toward a wider spread, but nevertheless, the opportunity appeared for personal research in the field of rating scales, and it is hoped that this unofficial and independent work may contribute something to those more intimately connected with the problem. The purpose of this project, then, is to construct a rating scale and, in the process, to make some useful contribution to the problem of rating non-commissioned officers of the first three pay grades in the U. S. Marine Corps.

---

1) The word "marking" will mean here the actual marks made by the reporting officer (used interchangeably with "rating officer") on the rating scale.

CHAPTER II

GENERAL CONSIDERATIONS IN THE CONSTRUCTION OF THE RATING SCALE.

A. Preliminary Considerations.

Specifically, it is believed by the writer that the primary use of a rating scale for NCO's of the U. S. Marine Corps is to determine their relative fitness for promotion, and the problem immediately resolves itself into one of meeting this requirement.[1] The problem was limited to the construction of a rating scale proper and therefore does not include a design for the entire fitness report, since the portion of the fitness report exclusive of the rating scale would seem to contain items which are properly matters for administrative policy determination.

The scale is not intended for use in evaluating candidates for enlistment but it is to be confined to the process of merit rating NCO's already in the service. Some secondary uses of the scale are: (1) to indicate general shortcomings in the calibre of NCO's in the Marine Corps so that plans for training or other remedial action can be initiated, and (2) to indicate where improvement might be made in the classification or recruitment policies.

---

1) An implication here is that there must be sufficient differentiation between individuals so that a choice can be made. This does not necessarily imply a normal distribution of ratings. For a discussion of this matter, see Hill, Robert L. Efficiency Ratings, Personnel Journal, 1937, 15, pp. 330-332.

To satisfy the main requirement, every reasonable means must be taken to cause raters to distribute their ratings so that they will be discriminatory. Methods that will tend to cause such a distribution are known and will be presented.

There are certain administrative problems which will make it desireable to keep statistical work at a minimum. With thousands of ratings per year to handle, the staff for this work could easily become too large for its proportionate share of the budget and make the use of the scale too costly. In the lower echelons, few, if any, trained statisticians or statistical machinery will be available. Practicability must be a major consideration in the construction of the scale.

The scale should be practical in another way too. It is possible that a new technique of rating-scale construction might prove to be a superior one, but it is the purpose of this project to construct a scale which could be put to immediate use. To increase the scale's chance for immediate practicability, no attempt will be made to create a new technique. The scale will be built on principles and techniques which are already operating satisfactorily.

It should be borne in mind that for service purposes a scale should measure proficiency rather than ability. There is an opportunity for the scale to be objective, since the ratings can be based upon observed performance and the scale constructed with that fact in mind. It would of course be little satisfaction for a commanding officer to have a group of NCO's of the highest ability who turned in poor performances.

In this instance, at least, the success of a rating scale should depend (aside from its ability to discriminate) on the reactions of two groups: the persons rating and the persons being rated. If the persons rated do not have confidence in the scale and understand it, its adverse influence on morale might more than offset its value. Likewise, if the raters who use the scale neither understand it nor want to use it, it will most certainly not be put to use in the way intended. All persons coming in contact with the scale must be sufficiently interested in it to master its values and limitations, for no scale yet devised approaches infallability.[1]

With these considerations in mind, the turn now is toward the more specific end of choosing a scale.

B. Types of Scales Considered.

Several types of scales were considered before a choice

---

1) An illustration is from Ewart, Seashore and Tiffin, A Factor Analysis of an Industrial Merit Rating Scale, Journal of Applied Psychology, 1941, vol. 25. The authors analyzed a twelve-trait scale using Thurstone's method and found that two factors accounted for nearly all of the total rating scores. Yet, even if a satisfactory two-factor scale were devised, its acceptance would be doubtful, for neither raters nor ratees in the Marine Corps would be likely to put much confidence in it.

was made. Such methods as the "paired comparison" and the "order of merit" or "ranking", though not strictly rating scales, were noted in passing and discarded as being too laborious as well as for being designed for small groups. Guilford gives some advantages of rating scales over the paired-comparison and rank-order methods; among them are: (1) rating scales require less time, (2) the procedure is more interesting to the rater, (3) rating-scale methods have a wider range of application; (4) ratings can be determined with psychologically naive judges who have had little training, and (5) rating scales can be used to rate large groups.[1]

The numerical rating scale is one consisting of descriptive phrases to which numerical or letter values are assigned. Some advantages of this type of scale are: (1) it is objective; (2) it is easy to score; (3) it is expedient to mark; (4) it is easy to construct; and (5) its form tends to make it unnecessary for a rater to maintain a subjective standard with relation to a large group. Some disadvantages are: (1) fine gradations in scoring are not feasible; (2) it is vulnerable to the "halo effect".[2]

---

1) Guilford, J. P. Psychometric Methods, P. 280.
2) The tendency for a rater to form a general impression of a person and rate him accordingly in a number of characteristics.

The man-to-man type of scale was formerly used by the Army. It has as its salient feature a "master scale" for each characteristic rated.[1] The rater selects individuals who are typical of the extremes and average of a characteristic and writes their names on his personal master scale. As individuals are rated they are compared to the master scale and rated accordingly. Some advantages of the man-to-man scale are: (1) the master scale is a theoretically stable standard; (2) it is easy to score; and (3) it is not difficult to construct. Disadvantages are: (1) it is relatively abstract; (2) it is vulnerable to the halo effect; and (3) in use it has proved too cumbersome.[2]

The defined-groups method is a method of rating persons with regard to their standing within a definite group, e.g. low, middle or high in relation to that group. Scoring may be simply numerical according to the subgroup in which the person rated is placed or, if a line is placed opposite each characteristic, may be the measured distance of a check mark from an end of the line. Some

---

1) The word "characteristic" will be used throughout in lieu of the word "trait" which is most frequently used in the literature on rating scales. The reason for this is that some of the items included in the scale constructed are not traits so that "characteristics" was chosen as a more inclusive term.

2) Scott, Clothier, Mathewson, and Spriegel, Personnel Management, p. 219.

advantages of this type are: (1) it is easy to mark; (2) it is easy to score; and (3) it is easy to construct. It has the following disadvantages: (1) it is abstract; (2) it is difficult for raters to keep in mind the degree of a characteristic in terms of symbols; and (3) it is vulnerable to the halo effect.

The check-list type of scale as advocated by Richardson and Kuder was considered.[1] In this scale there are no divisions by characteristics. A large number of statements, accurately scaled, are included, and the rater merely checks those applicable to the person rated. Advantages of this scale are: (1) it is objective; (2) it tends to minimize the halo effect; (3) it eliminates the necessity for a rater to maintain a subjective standard with relation to a large group; and (4) its reliability has been shown to be relatively high. Some of its disadvantages are: (1) its proper construction requires the cooperation of a large group; and (2) its application to

---

1) M. W. Richardson and G. F. Kuder, Making a Rating Scale that Measures, Personnel Journal, 1933, vol. 12, pp. 36-40.

rating characteristics is questionable.[1]

Army technicians have devised a new technique, the "forced-choice" scale. Numerous descriptive items are selected and scaled in a manner similar to the check-list type. It is generally characterized by a grouping of three of these items together and the raters having to make a choice for each item as to whether it describes the "ratee"[2] most, intermediate or least. Some advantages of the forced-choice scale are: (1) it is objective; (2) it tends to minimize the halo effect; and (3) it eliminates the necessity for a rater to maintain a subjective standard with relation to a large group. Some

---

1) A study was made of the monograph which would appear to be the father of the scientific construction of such scales. (L. L. Thurstone and E. J. Chase, The Measurements of Attitudes.) The writer cannot speak authoritatively, but he feels that there is not necessarily a logical transition from the measurement of attitudes using this method to the measurement of human characteristics. Thurstone's measurements would seem to be restricted to the measurement of attitude toward a single factor. Ewart, Seashore and Tiffin (A Factor Analysis of an Industrial Merit Rating Scale, Journal of Applied Psychology, 1941, vol. 25) found that at least two factors were present in a rating scale (termed "ability to do the present job" and "skill over and above the requirements for the particular job"). Most scales have more than two items to be rated. If we can assume two or more factors concerned in rating, then are these two factors properly weighted in the scientific check list? The items are nicely scaled, but in what proportion do they relate to the factors to be measured?

2) The word "ratee" is used here to mean a person for whom a proficiency rating is determined with the aid of a rating scale.

disadvantages are: (1) its proper construction requires the cooperation of a large group; and (2) complete data on its construction and use have not been publicized; and (3) raters have shown some resistance to it because it is difficult for a rater to tell how high or low he is rating.

The graphic rating scale has had the most widespread use of any rating scale. It is characterized by a straight line opposite each characteristic. Beneath the line are ranged phrases descriptive of varying degrees of proficiency in a characteristic. The rater may check anywhere along the line he wishes, and the rating is converted to a numerical score through the use of a stencil. Some advantages of the scale are: (1) the scale is simple and expedient for the rater to use; (2) it adapts to objectivity through the use of descriptive phrases; (3) it has had sufficient use to determine that it is well liked by many persons using it; (4) it tends toward a normal distribution of ratings since a rater can check anywhere along a line rather than on definite squares or symbols; (5) a great deal of recorded experience is available for use in the construction of such a scale; (6) its simplicity makes it easy to discuss a rating with a ratee; (7) its form tends to minimize the necessity for a rater to maintain a subjective standard with relation to a large group; and (3) it is relatively easy to construct. A disadvantage is that it is vulnerable to the halo effect.

Weighing the advantages and disadvantages of the various scales directed attention to three types: the check-list, the forced-choice, and the graphic. The main objection to the check-list and forced-choice types, from the standpoint of this work, was that the construction of such scales is no one-man job within anything like a reasonable length of time, so that both were somewhat reluctantly discarded.

The safest ground seemed to be in the direction of the graphic rating scale. It is a type which combines a degree of objectivity with fine gradations of scoring and yet remains relatively easy to construct.

The graphic scale is not a solution to all problems of rating. Available figures show its reliability to be somewhat less than that of the check-list type.[1] It can only minimize the "halo effect"[2] It does not eliminate variations in standards between raters but there are

---

1) H. E. Burtt, *Employment Psychology*, pp. 382-383.
2) W. V. Bingham, *Halo Valid and Invalid*, Journal of Applied Psychology, 1939, vol. 23, pp. 221-228 states that a certain amount of this effect may be desireable.

reasonably simple methods which tend to minimize the variations.

From the factors considered, the graphic type of scale was concluded to be the most practicable for this project.

C. Synthesis of Some of the Principles of Constructing a Graphic Rating Scale.

From the work of many persons, some established principles of rating-scale construction can be drawn. Many of these overlap and the trading of information between the experts has made the original authors somewhat obscure. Those listed below range from general rules to those applying particularly to the graphic scale. At least one source for each rule will be indicated; although in nearly every case, several writers on the subject of rating scales have substantiated each.

1. There is no one rating scale which is applicable to all situations.[1]

2. In determining the characteristics to be used in a rating scale, the logical procedure is to consult persons who are familiar with the occupations in question. The most frequently mentioned characteristics (in a survey) should be included in the final scale.[2]

---

1) H. E. Burtt, Employment Psychology, p. 350.
2) Ibid. p. 351.

3. Only those characteristics should be rated on a scale that are of greatest importance in the progress of an individual.[1]

4. Items in scales must be observable.[2]

5. In order to minimize halo effect, each characteristic should occupy a page by itself so that the rater can rate all his subordinates on a single characteristic before rating any of them on any other characteristic.[3,4]

6. Characteristics should be grouped according to the accuracy with which they can be rated.[5,6]

7. The characteristic upon which a person is to be rated should be a simple one - not a compound one.[7]

8. The characteristic should be defined and expressed objectively and not subjectively.[8,9]

9. The characteristic should be defined and expressed in terms of work performed on the job.[10,11]

---

1) Scott, Clothier, Mathewson, and Spriegel, Personnel Management, p. 232.
2) P. M. Symonds, Diagnosing Personality and Conduct, p.84
3) J. P. Guilford, Psychometric Methods, p. 271.
4) Scott, Clothier, Mathewson, and Spriegel, Personnel Management, p. 224.
5) J. P. Guilford, Psychometric Methods, p. 278.
6) P. M. Symonds, Diagnosing Personality and Conduct, pp. 104-108.
7) Scott, Clothier, Mathewson, and Spriegel, Personnel Management, p. 227.
8) Ibid, p. 229.
9) P. M. Symonds, Diagnosing Personality and Conduct, p.85
10) Scott, Clothier, Mathewson, and Spriegel, Personnel Management, p. 229.
11) H. E. Burtt, Employment Psychology, p. 357.

10. In order that the mind of the rater be restricted to the job, it may be desireable to omit the title of a characteristic but to describe it in terms of pertinent acts on the job.[1]

11. Only those characteristics that cannot be measured by an objective test should be subjectively evaluated on a rating scale.[2]

12. Only universally understood phrases should be used.[3]

13. Probably in ordinary work, the matter of weighting is unimportant.[4] However, it may be wise to assign rough weights where there are significant differences in the importance of characteristics.[5,6]

14. Ratings should be made in response to a pertinent question.[7,8]

15. The optimum number of scale divisions is five.[9]

16. The line should be of such a length that a stencil for scoring the rating can be easily calibrated.[10]

---

1) Scott, Clothier, Mathewson and Spriegel, Personnel Management, p. 229.
2) Ibid. p. 232.
3) P. M. Symonds, Diagnosing Personality and Conduct, pp. 66, 86.
4) Ibid, p. 83
5) H. E. Burtt, Employment Psychology, pp. 353, 354.
6) According to Tiffin and Musser, (Weighting Merit Rating Items, Journal of Applied Psychology, 1942, vol. 26, pp. 575-583) unless standard (z) scores are used, the items will weight themselves in a somewhat arbitrary manner.
7) Scott, Clothier, Mathewson and Spriegel, Personnel Management, p. 227.
8) J. P. Guilford, Psychometric Methods, p. 271.
9) Ibid. p. 263.
10) P. M. Symonds, Diagnosing Personality and Conduct, p.65

17. The line should be at least five inches long but not much longer so that it can be easily grasped as a whole.[1]

18. There should be no breaks or divisions in the rating line.[2]

19. Careful consideration of the descriptive phrases below the line is essential to the accurate use of the rating scale.[3]

20. Decide first on the phrases to represent the extremes of a characteristic.[4]

21. In describing characteristics, avoid the use of such general terms as "very", "extremely", "average", or "excellent".[5]

22. The end phrases of a scale should not be so extremely worded as to not be employed.[6]

23. The phrase descriptive of the neutral or average degree should be in the center of the scale.[7]

24. If there are five items, the intermediate ones should be closer in meaning to the center one than to the extremes.[8]

---

1) J. P. Guilford, *Psychometric Methods*, p. 271.
2) Ibid. P. 271.
3) Scott, Clothier, Mathewson and Spriegel, *Personnel Management*, p. 227.
4) P. M. Symonds, *Diagnosing Personality and Conduct*, p.65
5) J. P. Guilford, *Psychometric Methods*, p. 272.
6) Ibid. p. 271.
7) P. M. Symonds, *Diagnosing Personality and Conduct*, p.65
8) Ibid. p. 66.

25. The favorable extremes of a scale should be alternated in order to do away with a tendency to check at the margin of a page.[1,2]

26. Descriptive phrases should be set in small type, and there should be plenty of white space between them.[3]

27. The extreme phrases should be set flush with the end of the line.[4]

28. A finer distinction should not be required in rating than in scoring. If anything, the scoring unit may be smaller than the rating unit.[5]

29. In scoring, a stencil which divides each line into several sections to which numerical values are assigned should be used.[6]

30. The divisions of the scoring stencil need not be equal. They may be made to conform to the distribution of ratings.[7]

31. To encourage giving information not provided for on the forms, a space should be provided for that purpose.[8]

---

1) P. M. Symonds, *Diagnosing Personality and Conduct*, p.66
2) J. P. Guilford, *Psychometric Methods*, p. 271.
3) P. M. Symonds, *Diagnosing Personality and Conduct*, p.66.
4) J. P. Guilford, *Psychometric Methods*, p. 271.
5) Ibid. p. 272.
6) Ibid. p. 272.
7) Ibid. p. 272.
8) Scott, Clothier, Mathewson, and Spriegel, *Personnel Management*, p. 232.

32. The rater should be challenged (on the form) to illustrate the type of action that he regards as typical of the ratee.[1]

33. Judgments should be based on observations of concrete actions so that a space should be provided for checking "no opportunity to observe" or some similar expression.[2]

The principles listed above will serve as a guide in the construction of the scale for Marine non-commissioned officers.

D. Some Methods of Validating Rating Scales.

R. S. Driver lists some possible methods of establishing the validity of rating scales as follows:[3]

1. Comparison of the scale with some direct measure of performance.

2. Comparison with psychological tests purporting to measure the same ability.

3. Comparison with "work samples" (a short period of performance under controlled conditions.)

4. Analysis of the distribution of results. If ratings tend toward the normal curve, some investigators believe they are more valid.

---

1) Scott, Clothier, Mathewson, and Spriegel, Personnel Management, p. 227.
2) Ibid, p. 232.
3) R. S. Driver, The Validity and Reliability of Ratings, Personnel, Feb. 1941, pp. 185-191.

5. Analysis to determine the presence or absence of the "halo effect". Again some investigators believe that a normal distribution indicates absence of that effect and indicates a more valid scale.

6. Follow-up procedures - ratings compared with the progress of an individual in the organization.

7. Miscellaneous methods:

a. Comparison of rater's opinions with those of persons considered well informed.

b. Comparison of ratings with recommendations for salary changes.

c. Analysis of differences occurring in rating as a result of training raters.

d. Analysis of the consistency of ratings when an individual is rated in different departments.

e. Examination of the readiness with which a rater is willing to change his opinions as an indication of the value of the rating.

f. Comparison of the ratings of a group of individuals considered outstanding in the organization with a group of inferior individuals.

Most of the methods described above are of little use in this instance. No direct measure of performance will be available. The use of psychological tests as a criterion is questionable. A great many of them are validated with rating scales as criteria. Personality tests would necessarily be used to a large extent, and the methods of validating them are on somewhat shaky ground. The duties of NCO's in the Marine Corps are too diversified to lend themselves readily to "work samples".

If the opinion of Hill[1] is allowed credence, then a normal distribution of ratings would not necessarily be expected. The use of follow-up procedures would seem to be lifting one's self by one's bootstraps since ratings in the Corps should count heavily in determining the progress of an individual.

Some of the methods listed under seven above may be useful as general checks on a scale and will be given consideration.

One writer suggests that human judgments in themselves possess a degree of validity and suggests that the best method of achieving validity is in the initial selection of the characteristics.[2] One method is to get the opinions of several persons directly concerned or associated with the persons to be rated as to just which characteristics are important. The results of a tabulation and analysis of these judgments is later incorporated in the scale. The method will be followed in this project.

---

1) R. L. Hill, Efficiency Ratings, Personnel Journal, 1937, vol. 15, pp. 330-332.
2) P. M. Symonds, Diagnosing Personality and Conduct, p.108

E. Some Methods of Establishing the Reliability of Rating Scales.

While methods of determining the validity of a rating scale are not very successful, relibility can be determined with reasonable accuracy. There are two common methods. One is to correlate ratings of the same rater and ratees which are made at sufficient intervals so that the rater will not remember his previous ratings. The other method is to correlate the ratings of two or more different raters of the same ratees. Slawson concluded that two ratings by the same judge are no better than one rating.[1] Though this study has not been verified, it casts sufficient doubt on the first method to make the second one desireable where there is a choice. A reliability cooeficient of about .55 is typical under good conditions.[2] However, some scales have had a reliability as high as .87 considering the agreement of the rater with himself.[3]

F. Considerations in Instructing Raters.

A great deal of importance should be attached to the instructions accompanying the rating scale. Through these instructions those who have not had training in rating should be able to obtain some amount of specific information which will tend to increase the validity of their ratings. The raters' task should not be too tedious, and the instructions accordingly should be concise yet cover the salient points; they should be clear and free from technical language; they must fit the particular scale and purpose at hand; and they must make sufficiently obvious the pitfalls which the rater himself may be able to avoid and the undesireable effects which he may help to lessen.

---

[1] Quoted in P. M. Symonds, Diagnosing Personality and Conduct, p. 99.
[2] Ibid, p. 95.
[3] H. E. Burtt, Employment Psychology, p. 382.

It is convenient to divide the problem into two parts: (1) a set of specific instructions pertaining to the mechanics of marking the scale, and (2) a set of general instructions concerning errors and pitfalls in the subjective mental process of formulating ratings.

The formulation of the former will be delayed until the scale has been completed, but the considerations for the general instructions will be dealt with now, and later converted into the form in which they will be used.

It should be understood, first, that the success of any scale will depend largely upon the proficiency of the rater and that this proficiency can be improved by training.[1] There are many concepts of which raters need

---

1) J. H. Pockrass, Common Fallacies in Employee Ratings, The Personnel Journal, vol. 18, 1940, pp.262-267.

be aware. Most raters will be aware already of many of these, but a tangible confirmation of their own thinking is certainly desireable.

The rater should be aware of the primary objectives and uses of his ratings so that his interest and best effort will be aroused. He should thoroughly understand the importance of his work and know that it will be utilized.

The rating officer tends to biased and unfair ratings when judging friends, old acquaintances, or persons whom he particularly likes or dislikes.[1] Although knowledge of this fact cannot eliminate it, it should influence the rater to attempt to compensate for bias and attain an objective and impartial attitude.[2]

The rater must set and maintain a personal standard of rating. He should know, roughly, the normal distribution of scores to be expected when he is rating a group and should attempt to distribute his ratings discriminately rather than bunch them.[3]

Persons can be rated best after long and thorough personal observation by the rater, supplemented by notes recorded during the period of observation.[4] However, care

---

1) P. M. Symonds, Diagnosing Personality and Conduct, p.114
2) H. E. Burtt, Employment Psychology, p. 394.
3) Ibid, p. 395.
4) P. M. Symonds, Diagnosing Personality and Conduct, pp. 45-47.

must be taken that marks are based upon all the observed behavior and not just upon isolated good or bad examples. Too long acquaintance will tend toward higher marks for persons liked and lower marks for those disliked and should be guarded against by the rater.[1]

Some raters tend to rate too high and some too low. Others tend to rate all persons near the average. To help counteract these tendencies, the rater should know about what percentage of persons rated should fall within each division of a scale, but he can be told that his marks need not necessarily fall into a normal curve since he has a small sample.

The rater should be aware of the "halo effect", the tendency for the judge to form a general impression that is favorable or otherwise and to rate the person accordingly in a number of traits.[2] He can best minimize this effect by rating all the ratees on a single characteristic before rating any of the ratees on any other characteristic.[3]

There is a tendency to rate certain characteristics about the same regardless of the person concerned. The rater can help to offset this tendency by thinking objectively of each characteristic in terms of performance.[4]

---

1) H. E. Burtt, *Employment Psychology*, p. 393.
2) Ibid, p. 388.
3) Scott, Clothier, Mathewson, and Spriegel, *Personnel Management*, p. 224.
4) J. P. Guilford, *Psychometric Methods*, p. 275.

The best ratings will be from those characteristics of which the rater is sure of his judgment.[1] Certainty can be increased by requiring the rater to support his rating with short descriptions of conduct. These need not be given weight in the scale but can be used as a sort of motivation for objective thinking.

Raters find it easier to rate characteristics which are favorable than those which are unfavorable. For this reason, a rater might well spend more time in marking the unfavorable characteristics.[2]

For a fair and impartial rating to which corrections can be applied, the rater must do the rating unaided and neither ask for nor accept the advice of others in making the marks.[3]

Finally, the rater must allot sufficient time to do the rating.[4] If he understands the importance of his marks, he should not need coercion to take enough time. Careful and considered judgments are essential to fair rating.

These considerations are essential. On the basis of them, instructions to accompany the rating scale will be devised.

---

1) J. P. Guilford, *Psychometric Methods*, p. 277.
2) H. E. Burtt, *Employment Psychology*, p. 398.
3) P. M. Symonds, *Diagnosing Personality and Conduct*, p. 72
4) H. E. Burtt, *Employment Psychology*, p. 397.

CHAPTER III

THE MECHANICS OF CONSTRUCTING THE GRAPHIC RATING SCALE.

A. A Letter Questionnaire.

The first problem in actually constructing the scale was to choose the characteristics to be incorporated in the scale. It was decided to use opinions from a sample of Marine Corps officers. A form letter was devised (fig.1). It was written in business rather than official style since the project was of an unofficial character. The impression to be left with the recipients was that they need answer only if they were sufficiently interested to do so.

Rather than a check list of traits to be ranked, the letter consisted of a short explanation of the project and some specific directions for listing and defining characteristics. The purpose of framing the letter in that way was to avoid leading the officers to respond in a stereotyped way so that their suggestions would be their own and not influenced by prepared lists and definitions. As indicated in the figure, each addressee was requested to name and define the characteristics which he thought should be included in the scale and to indicate numerically the importance which he attached to each characteristic named.

The letter was sent to eighty-nine Marine officers, some by name and some by duties and station. Thirty-six

NAVAL R. O. T. C. UNIT
The Ohio State University
COLUMBUS 10, OHIO

Dear Sir:

The Marine Corps has ordered me to Ohio State University to do graduate work in Personnel Administration. In connection with this work, I am to do research consisting of the construction of a rating scale (fitness report) for Marine non-commissioned officers of the first three pay grades. It may be similar to the present NCO fitness report, or it may not, *depending largely on what you think.*

One of the first requirements for such a scale is the selection of items which are to be rated. The established method of doing this is to consult a group of persons who do the rating and combine items they consider essential in determining relative merit of NCO's.

I very much need from you, a list of characteristics or traits which you think should be rated in such a scale. Would you please just jot down on the back of this letter the items you consider important, with a short explanation of what you mean by each item? For example, one item might be *leadership*, a definition of which, but not necessarily yours, would be: *ability to elicit the cooperation of his colleagues and subordinates, to promote morale, and to develop a loyal and efficient organization.* Then would you indicate numerically the relative importance you attach to the first ten items, or the numerical rank of all the items if you have less than ten?

May I assure you that your name will not be used in connection with this work without your express permission. A self-addressed, postage free envelope is enclosed for your convenience, and I hope that I may hear from you soon.

If you should have any additional suggestions or comments, they will be greatly appreciated and certainly found useful in this project.

Thank you very much for your cooperation.

Very truly yours,

G. S. BAZE
Major, U. S. M. C.

FIGURE 1

TABLE I

REPLIES LENDING THEMSELVES TO TABULATION

RANK	NUMBER
Lieutenant Colonels	3
Majors	2
Captains	10
First Lieutenants	7
Unascertained	1
Total	23

TABLE II

REPLIES NOT LENDING THEMSELVES TO TABULATION

RANK	NUMBER
Lieutenant Colonels	3
Majors	2
Captains	6
First Lieutenants	1
Unascertained	1
Total	13

answers were received; of these thirty-six, twenty-three could be used for compilation, thirteen could not because no characteristics were listed. The sample was small, but a glance at Table 1 will indicate that it is a fair cross-section of a Marine officer population. A preponderance of the junior officers reported, but this is not believed to be undesireable since their lack of seniority is compensated by the fact that ordinarily company officers rate NCO's. It is possible that better results might have been obtained through a selective process, emphasizing to a greater extent the knowledge and experience of senior officers. However, this point could not be considered because addresses of senior officers were not available.

B. Tabulation of the Suggested Characteristics.

When the returns from the questionaire letter were received, the characteristics suggested were sorted according to definition. Definitions which were identical, or nearly so, were placed together. Thirty-one characteristics were isolated fairly well in this way. (Table 3).

The problem then was to determine which of these traits were considered sufficiently important to be included in the scale. An arbitrary method was used. The traits were tabulated as indicated in Table 3. The horizontal columns were numbered from one to thirteen (thirteen being

CHARACTERISTIC	RANK											
	1	2	3	4	5	6	7	8	9	10	11	12
1. Attention to Duty	5	0	1	0	3	0	0	0	0	1		
2. Attitude							1					
3. Ability	1											
4. Common Sense		1	1	6	3	1		1				
5. Cooperation	1	1	2	1	3	2	1	1	3	1	1	
6. Dependability			1	1								
7. Efficiency	1											
8. Force	1	1		1	1	2	2	3		2	1	
9. Handling Troops					1							
10. Initiative	2	3	2	2	1	2	2	1			1	
11. Instructing Ability								1				
12. Integrity			1	1			1					
13. Intelligence	2	4	1	4	1		1	1	1	1		
14. Knowledge of Job		1	3		1	1	1	1				
15. Leadership	5	1	3	5	1	1	2					
16. Loyalty (1)	3	3	1		2	2		2				1
17. Loyalty (2)	1											
18. Military Bearing	1											
19. Morals										1		
20. Obedience	1				1							
21. Past Record	1											
22. Perseverance		1			1	1	1		1			
23. Personal Appearance	1			2	3	4	4	1		1	1	
24. Physical Fitness			3		1		1	1	2	4		
25. Presence of Mind		3		1		1		1	2			
26. Proficiency in Rank	1											
27. Resourcefulness						1						
28. Sense of Responsibility		2			1							
29. Sobriety		1										
30. Progressiveness										1		
31. Open-Mindedness									1			

TABLE 3

Tabulation of Characteristics suggested by title and frequency with which a given rank was assigned.

the largest number of characteristics submitted by any one officer). Then the number of times a characteristic was assigned a particular rank was entered in the rows.

The relative importance of the characteristics was determined as follows. The one receiving the largest number of first rankings was deemed most important; the one receiving the largest number of first and second rankings (not counting the first characteristic) was deemed next in importance; the one receiving the largest number of firsts, seconds, and thirds (not counting the first two characteristics) was deemed of third importance; and so on. As indicated in Table 4, this procedure was halted after determining the relative importance of thirteen characteristics. The remaining eighteen were arbitrarily eliminated since each was suggested a maximum of three times and several of them only once. Another characteristic was eliminated from the list of thirteen. This was the last one, "persistence". The reason was two-fold: (1) persistence should have been combined originally with "attention to duty" since the definitions were extremely similar, and (2) it was discovered at this point that characteristics one through nine and eleven through thirteen corresponded to the characteristics listed in item nine of the NCO fitness report now in use. Item ten of Table 4 corresponded to the several characteristics in item eight of the NCO fitness report now used. A

	CHARACTERISTIC	RANK									
		1	2	3	4	5	6	7	8	9	10
1.	Leadership	5	(1)								
2.	Attention to Duty	5	0								
3.	Loyalty	(3)	3	1							
4.	Intelligence	2	4	1							
5.	Initiative	2	3	2							
6.	Common Sense	0	1	1	6						
7.	Cooperation	1	1	2	1	3					
8.	Personal Appearance	1			2	3	4				
9.	Force	1	1		1	1	2	2			
10.	Knowledge of Job		1	3		1	1	1	1		
11.	Physical Fitness			3		1		1	1	2	(4)
12.	Presence of Mind		3		1		1		1	2	0
13.	Perseverance		1			1	1	1		1	0

() around number indicates figure determining rank in case of tie.

TABLE 4

fairly obvious conclusion could be drawn: that the characteristics in item nine of the present NCO fitness report were considered in the main satisfactory by the Marine officers reporting.

The characteristics to be incorporated in the scale, then, were to be:

1. Leadership
2. Attention to duty
3. Loyalty
4. Intelligence
5. Initiative
6. Common sense
7. Cooperation
8. Personal appearance
9. Force
10. Knowledge of job
11. Physical fitness
12. Presence of mind

It should be borne in mind that, though the characteristics are listed by title for convenience, the important factor is the definitions behind them, regardless of what the title might be.

Doubtless there is overlap between the various characteristics. For example, one could argue reasonably that leadership is composed of any or all of the remaining characteristics. However, names for human characteristics seem limited only by the restraint and vocabulary of writers on the subject. For those characteristics termed "social" in particular, no rigid categories are developed and no certain method has been devised for isolating characteristics which are completely independent. Therefore, in the practical situation, it would seem more important to consider the raters' desires in the interest of gaining their cooperation. It appears to the writer that a primary consideration has been fulfilled: the characteristics included in the scale are those considered most important by the persons who do the rating.[1]

C. Definition of the Characteristics and Choice of Descriptive Phrases.

A good way to formulate definitions is to choose several definitions for each characteristic and submit them to a large number of judges for selection. For the descriptive phrases, choose a great many descriptive phrases for each trait. Then submit them to judges to arrange in rank order according to their strength from least to most. The average rank of each descriptive phrase then is its scale value and phrases equal units apart can be chosen for inclusion in the rating scale.

However, the method just described requires more staff and time than were available for this work so that a simpler method still was chosen. The method to be described lacks objectivity by comparison, but it is believed to be an adequate one, adding some validity to the construction of the scale.[2] Symonds states that the

---

1) See C. L. Shartle, Occupational Information, p. 126, for a brief discussion of this point.

2) H. E. Burtt, Employment Psychology, p. 373.

errors due to the unreliability of rating are greater than the errors due to carelessness of scaling, so that the latter can, for practical purposes, be disregarded.[1]

Characteristics would be tentatively defined and descriptive phrases tentatively selected. Then they would be attacked by a conference from four angles:

1. Are the definitions appropriate and objective?
2. Do the descriptive phrases fit the characteristic?
3. Are the descriptive phrases scaled correctly within each group?
4. Are the strengths of the descriptive phrase groups equal to each other?

Accordingly, in the light of the definitions suggested by Marine officers, the characteristics were tentatively defined and put in the form of questions. Five descriptive phrases were then tentatively selected for each characteristic. Several industrial and service scales were reviewed for ideas on phraseology, and Roget's Thesaurus and Colonel Munson's book on leadership[2] were consulted for the same purpose. The selected list follows (first draft):

---

1) P. M. Symonds, Diagnosing Personality and Conduct, p. 86.

2) L. H. Munson, Leadership for American Army Leaders.

1. Leadership: Has this NCO been successful in directing, controlling and influencing others, while maintaining high morale?
   - a. Inspiring, men will follow him anywhere.
   - b. Handles men well; commands respect.
   - c. Gets the job done.
   - d. Men have little confidence in him.
   - e. Continued friction with subordinates.
2. Attention to duty: Has this NCO applied himself to the duties of his job, day in and day out?
   - a. Energetically applies himself; is not afraid of doing too much.
   - b. Always on the job; feels his responsibility.
   - c. Does the work assigned to him.
   - d. Needs occassional prodding.
   - e. Lazy; needs constant urging.
3. Loyalty: Has this NCO rendered willing and faithful service at all times to his superiors, his subordinates and the unit?
   - a. Complete, unfailing, energetic support.
   - b. Actively supports his men; his superiors and his unit; a booster.
   - c. Usually defends his men, his superiors and his unit.
   - d. Loosely criticizes his men, his superiors and his unit.
   - e. Never misses a chance to knock, growl or gripe.

4. Intelligence: Has this NCO followed instructions easily, learned readily, and caught on quickly to new situations?

    a. Exceptionally quick.
    b. Learns with ease.
    c. Ordinary.
    d. Slow to learn.
    e. Dull.

5. Initiative: Has this NCO acted on his own responsibility in the absence of orders?

    a. Looks ahead; ready for proper action at right time.
    b. Quickly sees things he can do and gets them underway.
    c. Does most things right without being told.
    d. Sees things to do but asks first.
    e. Doesn't see things to do; gives up easily.

6. Common sense: Has this NCO considered all sides of problems presented to him and arrived quickly and certainly at the right solution?

    a. Comes up immediately with the right answer.
    b. Clear, quick thinker; usually right.
    c. More often right than wrong.
    d. Has trouble finding the right solutions.
    e. Seldom draws the right conclusions.

7. Cooperation: Has this NCO been successful in dealing with people? Has he been helpful and considerate in manner and act to his superiors, subordinates, fellow

NCO's and civilians?

a. Always helpful and considerate to everyone.

b. Usually has the feeling and interest of others in mind.

c. Conforms; pleasant to deal with.

d. Awkward in dealing with others.

e. Antagonizing.

8. Personal appearance: Has this NCO favorably impressed others by his physique, bearing and manner?

a. Commanding; excites admiration.

b. Noticeable; definitely an asset.

c. Impresses others favorably.

d. Impresses others negatively.

e. Sloppy.

9. Force: Has this NCO carried out his objectives with persistance and firmness?

a. Dynamic; commands obedience.

b. Pursues his goals with vigor and confidence.

c. Steers the middle course.

d. Retiring; lacks confidence.

e. Gives up quickly.

10. Knowledge of job: Has this NCO demonstrated mastery of the technical knowledge necessary in his job?

a. Knows all there is to know.

b. Has thorough working knowledge.

c. Knows enough for satisfactory performance.

d. Gets by with difficulty.

e. Doesn't know enough to perform efficiently.

11. Stamina: Has this NCO demonstrated physical stamina and endurance under hardship, adversity or discouragement?
    - a. Always ready for more; nothing gets him down.
    - b. Hard to get him down.
    - c. Physically fit.
    - d. Tires easily; lacks stamina.
    - e. Always tired; can't take it.
12. Presence of mind: Has this NCO shown that he can function promptly and effectively under pressure?
    - a. Super effective.
    - b. More clearheaded than ever.
    - c. Reliable and normal.
    - d. Less effective than usual.
    - e. Loses his head.

The list was then submitted to a conference comprised of:

1 Captain, U.S.N.
1 Commander, U.S.N.
1 Lt. Colonel, U.S.A.Reserve, (Professor of Political Science).
2 Lt. Commanders, U.S.N.
1 Lieutenant (S.G.) U.S.N.

The fact this conference was comprised predominantly of persons in the Naval Service should not be important since the judgments were of a general rather than a technical nature, and all have had experience in the armed services.

A separate conference was held with a Marine Captain and a Staff Sergeant, U. S. Marine Corps.

Extensive modifications were made. The principal ones were:

1. Loyalty was divided into two definitions, one for loyalty to superiors and one for loyalty to subordinates, since the tentative definition was considered "double-barrelled".

2. In two cases the number of descriptive phrases was reduced to four equally spaced phrases because the degree of possession of those particular characteristics seemed to fall more logically into such a grouping.

3. In one case, the number of descriptive phrases was reduced to four but the spacing was left intact. In this case, the conferees agreed that a blank was better than any descriptive phrase suggested.

4. In one case, the number of descriptive phrases was reduced to three. The argument here was that three degrees were as many as could be differentiated.

The revised definitions and phrases are listed below:

1. Has this NCO been successful in obtaining the spirited, cheerful, and willing obedience and cooperation of his men?
    - a. Men will follow him anywhere.
    - b. Handles men well.
    - c. Usually gets cooperation of men.
    - d. Men have little confidence in him.
    - e. Continual friction with subordinates.

2. Has this NCO applied himself to the duties of his job, day in and day out?

   a. Never spares himself.
   b. Does more than expected of him.
   c. Does the work assigned to him.
   d. Needs occassional prodding.
   e. Lazy; needs constant urging.

3. Has this NCO been loyal at all times to his superiors?

   a. Complete, unfailing, energetic support.
   b. Actively supports his superiors.
   c. Usually stands up for superiors.
   d. Carelessly criticizes superiors.
   e. Continually gripes about superiors.

4. Has this NCO been loyal at all times to his subordinates?

   a. Complete, unfailing, energetic support.
   b. Actively supports his men.
   c. Usually stands up for his men.
   d. Carelessly criticizes his men.
   e. Never says a good word about his men.

5. Has this NCO followed instructions accurately, learned readily, and adapted himself quickly to new situations?

   a. Exceptionally quick.
   b. Catches on easily.
   c. Ordinary.
   d. Slow to catch on.
   e. Dull.

6. Has this NCO performed effectively in the absence of specific orders?

    a. Prompt in taking appropriate action.
    b. Does most things right without being told.
    c. Asks before taking action.
    d. Waits to be told.

7. Has this NCO picked the right solution to problems confronting him?

    a. Comes up quickly with the right answer.
    b. Usually has good answer.
    c. More often right than wrong.
    d. Has trouble finding a good answer.
    e. Seldom right.

8. Has this NCO worked in harmony with people in the organization?

    a. Unusually helpful and considerate.
    b. Pleasant to deal with.
    c. Conforms.
    d. Irritates.
    e. Antagonizes.

9. Has this NCO favorably impressed others by his military appearance?

    a. Inspires admiration.
    b. Attracts attention.
    c. Unimpressive.
    d. Sloppy.

10. Has this NCO shown drive in carrying out his duties?

a. Dynamic.

b. Vigorous.

c. Active.

d. Blank.

e. Gives up easily.

11. Has this NCO demonstrated mastery of the technical knowledge necessary in his job?

a. Has complete knowledge.

b. Well grounded.

c. Knows enough for satisfactory performance.

d. Gets by with difficulty.

e. Insufficient knowledge.

12. Has this NCO shown physical endurance under difficult conditions?

a. Always ready for more.

b. Hard to get him down.

c. Physically fit.

d. Tires easily.

e. Can't take it.

13. Has this NCO shown that he can keep his head during an emergency?

a. More clearheaded than usual.

b. Reliable.

c. Loses his head.

D. Weighting the Items in the Scale.

Attention was turned next to weighting the various characteristics. It would have been feasible to determine weights from the data of Table 3, but the persons awarding ranks at that stage of the project had not seen the selected list of characteristics as a whole and this could make an appreciable change in their perspective. It seemed safer to submit the completed list and definitions for a new set of judgments. The scheme selected was to have judges distribute one thousand points among the items of the scale, allotting points to each characteristic according to the judges' opinion of its relative importance.[1] It was possible to obtain ten qualified judges:

1 Commander, U.S.N.
6 Lt. Colonels, U.S.A.
1 Major, U.S.M.C.
1 Captain, U.S.M.C.
1 Staff Sergeant, U.S.M.C.

The combined point totals for each item are shown in Table 5.

As previously noted, the problem of weighting is very controversial. But the points allotted to one characteristic were so much greater in number than the points allotted to the others that a rough tentative weight was assigned:

Leadership	2
All others	1

It was not believed desirable to make a finer distinction. The value of weighting items which are not to be converted to standard scores (as will be noted later) is doubtful, since the variability of the various distributions will alter the assigned weights. FINALLY, WEIGHTS WERE OMITTED.

---

[1] As suggested by P. M. Symonds, Diagnosing Personality and Conduct, p. 83.

COMBINED POINT TOTALS; TEN PERSONS EACH DISTRIBUTING 1000 POINTS AMONG THE CHARACTERISTICS LISTED BELOW:

Characteristic	Points
1. Leadership	1480
2. Attention to duty	820
3. Loyalty (up)	900
4. Loyalty (down)	910
5. Intelligence	740
6. Initiative	660
7. Common sense	630
8. Cooperation	660
9. Personal appearance	500
10. Force	670
11. Knowledge of job	780
12. Stamina	590
13. Presence of mind	660
	10,000

TABLE 5

E. The Form of the Rating Sheet.

The form of the rating sheets involved consideration of the following principles stated in Chapter II:

1. It was to be a scale adapted to rating single traits at a time.[1] Each sheet therefore needed to be identified with an NCO's name and rank, and the sheets for each of the various characteristics needed identifying numbers. A space for the last date of the reporting period was considered necessary in order to avoid confusion.

2. The scale line needed to be five inches long in order to be easy to visualize.

3. A blank would be provided for "no opportunity to observe".

4. Blanks would be provided on each sheet for specific illustrations of performance in the rated characteristic.

One rule was consciously violated in the form: The extreme descriptive phrases were not made flush with the ends of the lines. They were centered on the line ends so that half the phrase extended beyond the line. This procedure was followed because it was believed that most

---

1) It may be that from an administrative standpoint, single sheets would be impracticable. If so, the characteristics could all be placed on one page and the extreme phrases alternated to help compensate for halo. The latter method has demonstrated its practicability; however, it is believed that using single sheets is better from a scientific standpoint and not too impractical to be used.

raters would infer that the exact point on a line represented by the intermediate phrases would be the center of these phrases. To be consistent then, the extreme phrases would necessarily be centered on the ends of a line.

Appendix 1 contains samples of the finished rating sheet for each characteristic.

F. Formulation of Instructions to Accompany the Rating Scale.

General instructions for raters were written, based on the considerations stated in Chapter II.

It was believed desireable to:

1. Restrict the instructions to those directly concerning the persons performing the rating.

2. Present them as positively, personally and readably as possible.

3. Present them as a basis for further thought on the part of the rater.

Specific instructions for the mechanical marking of the scale were also written, and both general and specific instructions combined into a single set. The completed instructions are illustrated in Appendix 2.

G. Testing the Reliability of the Scale.

The rating scale was mimeographed and cut into sheets. Each sheet contained only one characteristic and was to be used for only one ratee (see Appendix 2). A copy of the instructions and sufficient forms to rate seven NCO's were given to two officers: a Marine Captain and a Navy Lieutenant (Senior Grade). The officers read the instructions and then rated a group of seven ratees well known to both of them. The ratee group consisted of one Staff Sergeant, U. S. Marine Corps and six Navy Petty Officers corresponding in rating to the rank of Marine NCO's of the first three pay grades.

This group of ratees was the largest that could be obtained at the writer's station, and it is unfortunate that there was not an all Marine group.

The forms were scored with a ruler marked in tenths of an inch, and for convenience the scores were recorded with the decimal place omitted. Tables 6 and 7 contain the tabulated scores for the group as rated by the two raters. Some scores were marked N for "no opportunity to observe". Correction was made for the items not observed by dividing the total score by the number of characteristics observed and multiplying this figure by thirteen (the total number of characteristics in the scale).

The corrected total scores of the two raters were then correlated using the "Product-moment" correlation formula.[1]

---

1) Peters and Van Voorhis, Statistical Procedures and Their Mathematical Bases, p. 97.

TABULATION OF SCORES OF FIRST RATER.

	Characteristic							
NCO	1	2	3	4	5	6	7	8
1	14	28	28	N	25	1	26	12
2	48	40	39	25	37	15	39	49
3	27	27	27	24	39	35	26	38
4	27	28	26	25	25	18	25	26
5	27	29	29	25	25	20	24	26
6	38	40	39	25	39	15	41	49
7	46	39	38	25	25	15	39	49

NCO	9	10	11	12	13	Total	Corrected Total
1	18	15	38	27	N	232	274
2	32	29	37	39	25	454	454
3	19	27	37	N	24	350	379
4	20	28	25	N	N	273	323
5	18	27	25	N	N	275	325
6	35	39	25	25	N	410	445
7	34	40	38	47	29	464	464

TABLE 6

TABULATION OF SCORES OF SECOND RATER.

NCO	Characteristic 1	2	3	4	5	6	7	8
1	20	19	20	20	36	29	35	18
2	39	30	19	34	37	49	39	49
3	29	26	24	N	32	19	N	37
4	16	33	30	N	17	36	18	30
5	27	26	29	N	21	2	23	31
6	33	38	37	N	43	48	35	46
7	37	33	20	N	37	N	31	36

NCO	9	10	11	12	13	Total	Corrected Total
1	29	23	39	N	21	309	335
2	44	39	41	31	15	466	466
3	27	29	25	N	N	248	358
4	26	22	33	14	27	302	325
5	35	26	19	32	25	296	321
6	29	35	11	37	27	419	455
7	36	27	37	34	N	328	426

TABLE 7

A correlation coeeficient of .92 was the result. The calculations are illustrated in Table 8.

Fisher's formula for the "most likely population equivalent of an "r"[1] was then applied and .91 obtained with the coeficient carried to two decimals (see Figure 2).

The probable accuracy of this coeficient was determined by computing "Student's "t" for the reliability of an "r" when estimated from the sample".[2] The t obtained was 4.91 (see Figure 3) This t was compared with Table XLV of Peters and Van Voorhis, Statistical Procedures and Their Mathematical Bases. Interpolation from the table indicated that .14% is the probability of obtaining an r of .91 by chance with the sample used; or odds of about 700 to 1. Fisher would term the reliability "highly significant".[3]

From the figures above, one would assume that the scale possessed a markedly high degree of reliability. However, it seems safe, because of the small size of the sample, to say only that the scale possesses some reliability. It is doubtful that a sample of less than ten persons would reveal the reliability of this scale with any accuracy.[4] For this

---

1) Peters and Van Voorhis, Statistical Procedures and Their Mathematical Bases, p. 153.
2) Ibid., p. 154.
3) Ibid., p. 176.
4) Ibid., p. 155.

CORRELATION OF RATING SCORES.

NCO	X	Y	x	y	x^2	y^2	xy
1	27.4	33.5	-10.7	-4.9	114.49	24.01	52.43
2	45.4	46.6	+ 7.3	+8.2	53.29	67.24	59.86
3	37.9	35.8	- .2	-2.6	.04	6.76	.52
4	32.3	32.5	- 5.8	-5.9	33.64	34.81	34.22
5	32.5	32.1	- 5.6	-6.3	31.36	39.69	35.28
6	44.5	45.5	+ 6.4	+7.1	40.96	50.41	45.44
7	46.4	42.6	+ 8.3	+4.2	68.89	17.64	34.86
TOTAL	266.4	268.6			342.67	240.56	262.61

Mean of X = 38.1 Standard Deviation of X = 6.93

Mean of Y = 38.4 Standard Deviation of Y = 5.87

$$\underline{r} = \frac{\Sigma xy}{N\,\sigma x\,\sigma y} = \frac{262.61}{284.75} = .92$$

TABLE 8

Most likely value of r

$$\hat{p} = r - \frac{r(1-r^2)}{2(N-1)} \left[1 - \frac{1-5r^2}{4(N-1)} \right]$$

$$= .92 - \frac{.92(1-.8464)}{12} \left[1 - \frac{1-4.232}{24} \right]$$

$$= .92 - .01336$$

$$= .91$$

FIGURE 2

Estimate of probability of obtaining by chance an r of the size obtained

$$t = \sqrt{\frac{r^2(N-2)}{1-r^2}} = \sqrt{\frac{5 \times .8281}{1 - .8281}} = 4.91$$

FIGURE 3

reason, any further statistical analysis was reserved until such time as a larger sample could be obtained.

CHAPTER IV

SOME METHODS OF COMPENSATING FOR VARIATIONS IN RATING SCALES.

It can safely be said that no method of compensating for variations in raters' standards can be completely satisfactory, but that some method is essential to get ratings on a comparable basis. Schemes varying from the very simple to the extraordinarily complex have been devised. Some of the simpler ones are adequate, luckily, for the more complex and long statistical processes are not practical to correct a scale used to rate a large number of persons. Some elaborate systems are suggested by Guilford[1] and Bent.[2] They have in common the unfortunate characteristic of being impractical for large groups. Chosen for consideration in connection with this rating scale are three less complicated methods:

1. Pooled ratings
2. Standard scores
3. Letter grades

A. The Method of Pooled Ratings.

"The objection that ratings are subject to many constant errors is met by proof to the contrary that pooled ratings somehow eliminate the force of these errors."[3]

---

[1] J. P. Guilford, *Psychometric Methods*, p. 272-274.
[2] R. K. Bent, *Various Techniques of Combining Ratings*, Journal of Educational Psychology, 25, pp. 65-70.
[3] J. P. Guilford, *Psychometric Methods*, p. 280.

The pooling of ratings simply means having more than one rater rate each ratee. The difficulty is in getting a sufficient number of raters who know the ratee well enough to give a fair rating. Symonds says that the pooled judgments of five to ten raters are necessary to get sufficient reliability.[1]

Others have drawn the conclusion that three or four ratings averaged will locate the person in the scale with sufficient accuracy.[2] Arithmetic averages of the ratings would then provide a fair composite score on each occassion a rating is made. Aronson noted that in the U. S. Civil Service even three qualified raters usually cannot be found.[3] That may well be the case in the Marine Corps.

A more momentous objection to the method is that the opinions of the more senior officers might be reflected in their juniors'. However, the tendency could probably be offset by not allowing any one rater to see the marks made by another, or by having junior officers rate first.

B. The Method of Standard Scores.

A standard score is the deviation of a particular score from the mean of a distribution of scores divided by the

---

1) P. M. Symonds, Diagnosing Personality and Conduct, p. 96
2) Scott, Clothier, Mathewson and Spriegel, Personnel Management, p. 233.
3) A. H. Aronson, Service Rating Plans, Public Personnel Review, 1941, 1, 2, pp. 298-305.

standard deviation of the distribution. The scores are thus reduced to common units and can be compared with one another.[1]

To compute standard scores, continuous records would be kept of a rating officer's scores for all the NCO's he might mark. The mean and standard deviation for a particular rating officer would be computed. The deviations from the mean for each NCO rated would be found and divided by the standard deviation to compute his standard score.

All standard scores tend to range between plus and minus three. To get the scores on a more familiar basis, they could be multiplied by ten and seventy added to each. Then the range would be between plus 40 and 100.

Though the statistical procedures for computing the standard scores of one distribution is not exceedingly laborious, when multiplied by a thousand or more, the mechanics of computing them, say twice yearly, would necessitate a large group of computers and all skilled at operating a calculator, not to mention the acquisition of the calculating machines. In addition, the method was intended to correct more formal tests, and its accuracy is somewhat wasted in correcting scores derived from opinions.

C. The Use of Letter Grades.

A simpler and more convenient method described by Burtt

---

1) Peters and Van Voorhis', Statistical Methods and Their Mathematical Bases, p. 80.

seems to the writer to be adequate:

"A scheme that has sometimes been used consists of taking a considerable number, perhaps 50, of ratings made by a given individual, arranging them in order from best to worst, and calling the best 10 percent A, the next 20 percent B, the middle 40 percent C, the next 20 percent D and the lowest 10 percent E. Thus, after all, whether the rater takes a high or low standard, those whom he puts relatively high will receive an A rating, indicating desireable possession of the trait, and those with relatively low ratings will receive a grade of E."[1]

The labor in applying this scheme would be considerably reduced from that of the standard scores method. One difficulty is that the scheme could not be applied until a rater had rated the prescribed number of NCO's; another is that the system might be unfair when the NCO's rated by a particular officer really were within a very narrow range of proficiency. But these difficulties are likewise encountered using standard scores.

The problem finally resolves itself into whether any system of correction should be used. In the opinion of the writer, the advantages of using some system unquestionably outweigh the disadvantages, and the most expedient system of correcting ratings of NCO's is through use of the letter-grade system.

---

1) H. E. Burtt, Employment Psychology, p. 379.

CHAPTER V.

A SCHEME FOR THE USE OF THE SCALE.

The construction of the scale being complete and systems for correcting ratings having been considered, the problem remains of putting these things into a workable arrangement. How should the rating scale be incorporated into the total fitness report, and what scheme should be used to put rating corrections into effect?

A. Incorporating the Scale into the Total Fitness Report.

To incorporate the scale into the fitness report would necessitate two changes in the report. First, items six and seven, which are instructions to raters, must be replaced by the set of instructions applicable to the new scale. Then, the new scale must be substituted for items eight and nine, which contain the scale used at present.

The form of the scale here constructed dictates that it be separate from the rest of the report. A solution is to print the fitness report (less the rating scale) on an envelope, size 8½ by 11 inches. The rating sheets could be printed on thirteen separate pads, one characteristic to a pad, with perforations between the rating sheets so that each could be removed singly.

Then after rating characteristics one at a time, the rater could sort the rating sheets into the envelopes according to NCO's rated and forward them.

The raters' instructions would also be separate from the report forms, but distributed along with and in proportion to the report forms and rating sheets.

B. Correcting Variations between Raters.

Two copies of the report would be forwarded to Headquarters. One copy would be filed in the reporting officer's file; the other copy would be filed in the NCO's file. As the reports are placed in the officer's file, they could be numbered. The file clerk would in that way be notified automatically when enough reports were in an officer's file so that corrections to his reports could be made. This number could be fifty or less. The exact number could be determined arbitrarily by observation of a sample of all officers' reports to see whether range limits of rating scores were becoming firm.

Then a range for each reporting officer could be established between his high and low scores. The upper 10% of his range could be given a grade A, the next 20% B, the next 40% C, the next 20% D, and the last 10% E.

Grades would then be transferred to the NCO's files. Ranges could be checked periodically as determined from experience.

The method of ranges is recommended in preference to arranging the scores in order from best to worst as described in Chapter IV, C, because it is quicker and easier when large numbers of reports are dealt with.

No doubt many minor problems would arise in connection with the administration of this scheme, but there is no apparent reason why they could not be easily solved as they arise.

CHAPTER VI.

CONCLUSIONS.

It is believed that the scale may be used effectively in its present form. However, certain improvements should be made in the scale after it has been used by a larger number of raters to rate a larger number of NCO's. Among these improvements are:

A. The reliability of the scale should be checked on a larger sample.

B. Scores on single characteristics could be intercorrelated with (1) total scores and (2) independent judgments of an NCO's general efficiency with a view to determining whether a lesser number of characteristics would yield a score close to that of the thirteen now incorporated in the scale. Such a procedure would probably result in future revision to a shorter and therefore simpler scale. It would also contribute to its validity by eliminating the less discriminatory characteristics.

C. More logical weights for the remaining characteristics might be determined arbitrarily on the basis of the intercorrelations.

D. Improvements in the definitions and descriptive phrases no doubt can be made and should be from time to time. There is no reason for the scale to become static. One of the advantages of the graphic scale is that it lends itself to change with a minimum of disruption of administration.

E. Thought might be given to the possibility of making scoring stencils calibrated so that there would be a stencil to fit (approximately) any rater's distribution of marks. Then corrected scores, or even letter grades, could be read directly from the stencil as the scores are measured.

F. Until sufficient reports have been received to approximate a normal distribution, it probably would be desireable to assume a mean and standard deviation for all raters and make corrections of standards on that basis.

G. Thought might also be given to determining a percentage correction for each rating officer, based on his range as compared with the average range, for use in lieu of a system such as that described in E above.

H. In constructing the graphic scale, the writer has formulated a suggested procedure for further research, using the graphic scale as a basis, provided the close cooperation of, say, 50 Marine officers could be obtained; these are:

1. Intercorrelate the scores on characteristics listed in the graphic scale as suggested in paragraph B above.

2. Eliminate any characteristics whose correlations with total scores are very low and whose correlations with other characteristics are very high.

3. Construct a modified check-list type of scale

(modified in that items would be scaled under each characteristic rather than scaled in a heterogeneous mass) using Likert's scaling technique[1] or a similar one.

4. Draft the scale with the ~~titles~~ ITEMS mixed and the characteristic titles and definitions omitted.

A large emphasis has been placed on correcting raters' standards, and the writer wishes to emphasize this aspect once more and say that, in his opinion, corrections are essential if the ratings of different raters are to approach comparability.

---

1) R. Likert, _A Technique for the Measurement of Attitudes_.

APPENDIX 1

INSTRUCTIONS FOR RATERS

The usefulness of the marks you make on a rating scale depends entirely on you. The importance of these marks cannot be overestimated. The completed fitness report will be used for:

1. Determining the order of promotion of NCO's.
2. Determining the caliber of NCO's in the Corps.
3. Indicating where improvement may be made in classification or recruitment policies.

This means that you will not only determine who will be eligible for promotion but that you will -- to a large extent -- determine the quality of NCO's in the Corps as a whole; consequently, your markings must be made as truly, carefully and accurately as possible.

The type of rating scale used with the fitness report form is called the "Graphic Rating Scale". It is characterized by a line for each characteristic being rated; under which are arranged phrases descriptive of the characteristic possessed by an NCO. The mechanics of marking are set out at the end of these instructions. Here are some brief pointers, which, if kept in mind, should be helpful while you are in the process of marking the scale:

1. Allot yourself a definite time to do the rating and take plenty of it. Careful and considered judgment

is essential to fair rating.

2. Do it all yourself. Neither ask for nor accept advice of others in making marks. You cannot be consistent if you do.

3. It is easier to rate the good characteristics than the bad, so take more time on the bad ones.

4. Your marks are likely to be biased one way or the other when you are rating friends, old acquaintances, or persons whom you greatly like or dislike. Be as open-minded as you can and base your marks on actual performance.

5. Some officers tend to rate consistently high or low. You must set and maintain your own personal standard of rating. To do this, keep in mind that normally your rating will be distributed like this:

Highest fifth	10%
Next highest fifth	20%
Middle fifth	40%
Next lowest fifth	20%
Lowest fifth	10%

6. Your best ratings of NCO's will come from long and thorough personal observations supplemented by notes you have taken during the observation period. But be sure that you record representative performance and not just outstanding good or bad examples; and watch your tendency to higher marks based solely on a very long acquaintance.

7. You may have a tendency to rate certain characteristics about the same regardless of the persons involved. You can help to offset this tendency by thinking objectively of each characteristic in terms of performance.

8. The "halo effect" is the very common tendency to form a general impression of a person and rate him accordingly in a number of characteristics. This effect is minimized by rating all persons on one characteristic before going to the next characteristic.

9. Your best rating will be those of which you are sure and that is why you are asked to write in after each characteristic a very short description of representative performance. In addition this will help you to pin down the performance for yourself.

10. Be sure to re-read these instructions each time you have fitness reports to make and devise additional pointers of your own which will help you to make a fair, accurate and impartial rating.

11. These are the directions for marking the scale:

a. For each numbered item, there is a question which indicates a characteristic which an NCO should possess. Underneath each question is a line representing the quantity of this characteristic an NCO may possess; at one end is the lowest quantity and at the other end is the highest quantity. In between the two ends of the line lie the various degrees of a characteristic. Underneath the line are phrases descriptive of the various degrees. These phrases are to be used as a guide.

b. To indicate the degree or quantity of a characteristic which you believe an NCO to possess, place a check mark at the appropriate spot on the line. The check may be anywhere on the line. You do not have to place the check mark over a descriptive phrase.

c. Work through one characteristic for all NCO's before proceeding to the next characteristic.

d. Write in the space provided on the rating sheet a very brief sample of the performance on which you base your judgment.

e. If you have not observed a given NCO's performance in a given characteristic, place a check mark in the blank following "no opportunity to observe". Do not hesitate to use this blank if you have not had sufficient observation to form a judgment.

f. After you have worked through all the characteristics, sort them by NCO's and place the rating sheets for each NCO in his separate fitness report envelope.

g. When the fitness report envelopes and rating sheets are all correctly completed, forward them in accordance with current instructions.

APPENDIX 2

1.

Name: #3 Rank: ____________
Last First Middle Initial

Period Ending ____________
Day Month Year

Has this NCO been successful in obtaining the spirited, cheerful, and willing obedience and cooperation of his men?

Men will follow him anywhere	Handles men well	[✓ Usually gets cooperation of men]	Men have little confidence in him	Continual friction with subordinates

No opportunity to observe ____________

Give specific illustrations of performance on which your judgment is based ____________

2.

Name: #1 Rank: ____________
Last First Middle Initial

Period ending: ____________
Day Month Year

Has this NCO applied himself to the duties of his job, day in and day out?

Never Spares himself	Does more than expected of him	[✓ Does the work assigned to him]	Needs occassional prodding	Lazy: needs constant urging

No opportunity to observe: ____________

Give specific illustrations of performance on which your judgement is based: grudgingly at times will do more than is expected but only when results are beneficial to himself

3.

Name: ______________________ Rank: ____________

Period Ending ______________________
Day Month Year

Has this NCO been loyal at all times to his superiors?

Complete, unfailing energetic support	Actively supports his superiors	Usually stands up for superiors	Carelessly criticizes superiors	Continually gripes about superiors

No opportunity to observe: ______________________

Give specific illustrations of performance on which your judgment is based: ______________________

____

____

4.

Name: ______________________ Rank: ____________
Last First Middle Initial

Period ending: ______________________
Day Month Year

Has this NCO been loyal at all times to his subordinates?

Never says a good word about his men	Carelessly criticizes his men	Usually stands up for his men	Actively supports his men	Complete, unfailing energetic, support

No opportunity to observe ______________

Give specific illustrations of performance on which your judgement is based: ______________________

____

5.

Name: ______________________________ Rank: ______________
Last First Middle Initial

Period ending: ______________________
Day Month Year

Has this NCO followed instructions accurately, learned readily, and adapted himself quickly to new situations?

Exceptionally quick	Catches on easily	Ordinary	Slow to catch on	Dull

No opportunity to observe: ______________

Give specific illustrations of performance on which your judgement is based: ______________________________

____

6.

Name: #4 ______________________________ Rank: ______________
Last First Middle Initial

Period ending: ______________________
Day Month Year

Has this NCO performed effectively in the absence of specific orders?

Waits to be told	[✓ Asks before taking action]	Does most things right without being told	Prompt in taking appropriate action

No opportunity to observe: ______________

Give specific illustrations of performance on which your judgement is based: ______________________________

____

7.

Name: ______________________________ Rank: ______________
Last First Middle Initial

Period ending: ____________________
Day Month Year

Has this NCO picked the right solution to problems confronting him?

Comes up quickly with right answer	Usually has good answer	More often right than wrong	Has trouble finding a good answer	Seldom right

No opportunity to observe: ______________

Give specific illustrations of performance on which your judgement is based: ______________________________

____

8.

Name: ______________________________ Rank: ______________
Last First Middle Initial

Period ending: ____________________
Day Month Year

Has this NCO worked in harmony with people in the organization?

Antagonizes	Irritates	Conforms	Pleasant to deal with	Unusually helpful and considerate

No opportunity to observe: ______________

Give specific illustrations of performance on which your judgement is based: ______________________________

9

Name: ______________________________ Rank: __________
Last First Middle Initial

Period ending: ____________________
Day Month Year

Has this NCO favorably impressed others by his military appearance?

Inspires Admiration	Attracts Attention	Unimpressive	Sloppy

No opportunity to observe: ______________

Give specific illustrations on which your judgement is based: ______________________________

____

____

10

Name: ______________________________ Rank: __________
Last First Initial

Period ending: ____________________
Day Month Year

Has this NCO shown drive in carrying out his duties?

Dynamic	Vigorous	Active	Gives up easily

No opportunity to observe: ______________

Give specific illustrations of performance on which your judgment is based: ______________________________

____

____

11

Name:____________________________________ Rank:____________
Last First Middle Initial

Period ending________________________
Day Month Year

Has this NCO demonstrated mastery of the technical knowledge necessary in his job?

Insufficient knowledge	Gets by with difficulty	Knows enough for satisfactory performance	Well grounded	Has complete knowledge

No opportunity to observe:________________

Give specific illustrations of performance on which your judgement is based:____

____

12

Name:____________________________________ Rank:________________
Last First Middle Initial

Period Ending:_______________________________
Day Month Year

Has this NCO shown physical endurance under difficult conditions?

Always ready for more	Hard to get him down	Physically fit	Tires easily	Can't take it

No opportunity to observe:_____________________

Give specific illustrations of performance on which your judgement is based:____

____

____

Name: ______________________________ Rank: ________________
Last First Middle Initial

Period ending: ______________________
Day Month Year

Has this NCO shown that he can keep his head during emergency?

More clearheaded than usual ________ Reliable ________ Loses his head

No opportunity to observe: ______________

Give specific illustrations of performance on which your judgement is based: ______________________________

____

____

BIBLIOGRAPHY

BOOKS:

Bingham, W. V. *Procedures in Employment Psychology*, A. W. Shaw Co., Chicago, 1926.

Burtt, H. E. *Principles of Employment Psychology*, Harpers, New York and London, 1942.

Green, H. H. *Measurements of Human Behavior*, Odyssey Press.

Guilford, J. P. *Fundamental Statistics in Psychology and Education*, McGraw-Hill, New York and London, 1939.

Guilford, J. P. *Psychometric Methods*, McGraw-Hill, New York and London, 1936.

Mosher and Kingsley, *Public Personnel Administration*, Harpers, New York and London, 1936.

Mumson, E. L. Jr. *Leadership for American Army Leaders*, The Infantry Journal, Washington, D.C. 1941.

Peters and Van Voorhis, *Statistical Procedures and Their Mathematical Bases*, McGraw-Hill, New York and London, 1940.

Scott, Clothier, Mathewson and Spriegel, *Personnel Management*, McGraw-Hill, New York and London 1941.

Shartle, C. L. *Occupational Information*, Prentice-Hall, Inc New York, 1946.

Symonds, P. M. *Diagnosing Personality and Conduct*, The Century Company, New York and London, 1932.

PAMPHLETS AND PERIODICALS:

Aronson, A. H. *Service Rating Plans*, Public Personnel Review, 1941, vol. 2, pp. 298-306.

Bent, R. K. *Various Techniques of Combining Ratings*, Journal of Educational Psychology, vol. 28, pp. 65-70.

Bingham, W. V. *Halo Valid and Invalid*, Journal of Applied Psychology, vol. 23, 1939, pp. 221-228.

Driver, R. S. *The Validity and Reliability of Ratings*, Personnel, Feb. 1941, pp. 185-191.

Ewart, Seashore and Tiffin, *A Factor Analysis of an Industrial Rating Scale*, Journal of Applied Psychology, vol. 25, 1941.

Guilford, J. P. and Jorgensen, A. P. *Some Constant Errors in Rating*, Journal of Experimental Psychology, vol. 22, 1938, pp. 43-57.

Guilford, J. P. *The Computation of Psychological Judgments in Absolute Catagories*, Journal of Experimental Psychology, vol. 22, 1938, pp. 35-42.

Hill, R. L. *Efficiency Ratings*, Personnel Journal, vol. 15, 1937, pp. 330-332.

Likert, Rensis *A Technique for the Measurement of Attitudes*, Archives of Psychology Monograph, Number 140, June, 1932.

Lynch, J. M. *The Psychology of the Rating Scale*, Education Administration and Supervision, vol. 30, 1944, pp. 497-501.

Pockrass, J. H. *Rating Training and Experience in Merit System Selection*, Public Personnel Review, vol. 2, 1941, pp. 211-222.

Pockrass, J. H. *Common Fallacies in Employee Ratings*, Personnel Journal, vol. 18, 1940, pp. 262-267.

Remmers, H. H. *Further Studies in Attitudes*, Studies in Higher Education, Number 31, Purdue University, 1936.

Richardson and Kuder, *Making a Rating Scale that Measures*, Personnel Journal, vol. 12, 1933, pp. 36-40.

Rugg, H. O. *Is the Rating of Human Character Practicable?*, Journal of Educational Psychology vol. 12, 1921, pp. 425-438, 485-501; vol. 13, 1922, pp. 30-42, 81-93.

Smith, H. H. *The Santa Barbara Behavior Rating Scale*, Journal of Educational Research, vol. 37, 1944, pp. 500-511.

Smith, R. E. *Foreman Selection through Merit Rating*, Personnel, March, 1944, pp. 270-277.

Thurstone and Chave, *The Measurement of Attitudes*, University of Chicago Press, 1929.

Tiffin and Musser, *Weighting Merit Rating Items*, Journal of Applied Psychology, 1942, pp. 575-583.

War Department, Adjutant General's Office, Personnel Bureau, Personnel Research Branch (Restricted): *Construction and Scoring of the Officer Efficiency Report* OER-A, 10, October, 1945.

Comparison of the Rating Check List (RCL) and Forced Choice List (FCL) Methods of Obtaining Ratings, 9 July, 1946.

A Field Study of the Effectiveness of FCL-3a, A Self-Training Indorsed Efficiency Report, PRS Report No. 674.

Construction and Scoring of the Officer Efficiency Reports, FCL-2a, b and c, 9 October, 1945.

Construction and Field Evaluation of the Officer Efficiency Report, Form ECE-1, 21, January, 1947.

Validation of the Officer Efficiency Report, Form ECE-1, 22 January, 1947.

Major Study of Comparative Validity of Five Periodic Officer Efficiency Reporting Methods, 5 December, 1945.

Thesis
B25 6352

Baze
A rating scale

DE 24 58
AG 25 64 10014
14030

2

ACCOPRESS BIN
BG 2507 EMB
MADE BY
ACCO PRODUCTS, INC
LONG ISLAND CITY, N.Y., U.S.A.

thesB25
A rating scale for marine non-commission
3 2768 002 01537 2
DUDLEY KNOX LIBRARY